

Radio Resource Management Strategies for DVB-S2 Systems Operated with Flexible Satellite Payloads

Giuseppe Cocco*, Tomaso De Cola*, Martina Angelone[†], Zoltan Katona*

*German Aerospace Center (DLR), Institute of Communications and Navigation
82234, Oberpfaffenhofen, Germany

e-mail: {giuseppe.cocco, tomaso.decola, zoltan.katona}@dlr.de

[†]Communications and TT&C Systems and Techniques Section
ESA/ESTEC, Noordwijk, The Netherlands

e-mail: martina.angelone@esa.int

Abstract—The increasing demand for high-rate broadcast and multicast services over satellite networks has pushed for the development of High Throughput Satellite (HTS) characterized by a large number of beams (e.g., more than 100). Moreover, the variable distribution of data traffic across beams and over time has called for the design of a new generation of satellite payloads, able to flexibly allocate bandwidth and power. In this context, this paper explores the technical challenges related to radio resource allocation in the forward link of multibeam satellite networks and proposes a strategy based on a modified version of the simulated annealing algorithm and a newly proposed objective function to meet as close as possible the requested traffic across the beams while taking fairness into account. Performance results confirm the effectiveness of the proposed approach and also shed some light on possible payload design implications.

I. INTRODUCTION

The advent of High Throughput Satellite (HTS) systems [1], [2] has revolutionized the concept of satellite communications in that new systems operating in the Ka frequency band (and above) are being designed in order to provide geographical coverage through a large number of beams. Such dramatic change has started upon the ever-increasing user demand for broadcast/multicast services characterized by high rates and reliability performance. To meet these requirements, a natural technology candidate is the Digital Video Broadcasting - Satellite - Second Generation (DVB-S2) standard [3], which is nowadays one of the most widespread and preferred options from broadcasters of satellite systems in the forward link.

In spite of the attractive performance figures that can be attained by DVB-S2 (e.g., in terms of spectral efficiency), the problem of optimally allocating bandwidth to beams and optimally operate the payload from a power perspective according to the amount of requested traffic is still not completely solved. This is because of the large number of variables that play a role in the resulting radio resource allocation problem. Traditionally, this problem has been often addressed from a ground segment viewpoint, by proposing optimization frameworks able to take into account propagation impairments (e.g., rain) and interference contribution from other beams (e.g., co-channel interference (CCI)). For instance, reference [4] addresses the problem from a scheduling viewpoint, allocating different ModCods to the satellite beams. However,

the complex characteristics of data traffic (time- and space-correlation, heavily depending on the specific geographic area) have always represented a formidable obstacle against deriving closed-form solutions, hence requiring the introduction of approximated models or the use of numerical optimization techniques.

On the other hand, the recent years have also witnessed an evolution of satellite system concepts from a space segment viewpoint, which also have an important impact on the resource allocation problem [5]. Specifically, more sophisticated payload designs have been introduced [6], so as to cope with the time and geographic variations of the bandwidth requested by each beam. This mainly resulted in two possible design options, namely flexible and beam-hopping payloads. The latter makes use of a time-slotted illumination window so that it is possible to define the sequence of beam illumination and the number of slots assigned to each beam according to the traffic demands and the antenna radiation pattern. The former makes use of a dual approach, consisting in allocating bandwidth or power to beams in relation to the offered and requested traffic. Both options have obviously important implications on the specific payload design (e.g., number of traveling-wave tube amplifiers (TWTA) and structure of the payload connection matrix) and the related constraints (e.g., mass and available power) imposed by the technology available nowadays. In [7], [8] the problem of time/beam allocation is studied in presence of traffic asymmetry. In the paper a closed form solution for the optimal resource allocation in a simplified setup with no interference is derived for two different utility functions, aiming at matching the requested bitrate and maximizing the product of the ratios between the offered and requested capacity across the beams, respectively.

In [8], [9] the advantages of multi-beam with respect to single beam satellite systems is studied under different performance metrics. Specifically, the optimal power allocation is derived for two different objective functions, one leading to throughput maximization and the other related to fairness. Although the aforementioned papers offer interesting hints on the problem of resource allocation, the validity of the results is limited by the assumption of no co-channel interference, which is instead removed in [10]. In the paper a phased array antenna

is assumed at the satellite and call-admission control schemes are investigated. Differently from the approaches adopted in the papers mentioned above, the studies contained in [11], [12] explore the benefits of power allocation. In particular, a two-stages sub-optimal algorithm is applied to solve a non-convex optimization problem, whose solution offers some insights about the relations between power allocation and offered traffic on the forward link of satellite networks. Finally, beam-hopping and flexible systems are compared in [13], where the latter implement non-uniform bandwidth allocation and make use of sizable beams. It is worth noting that most of optimization strategies considered in the available literature with respect to resource allocation make use of genetic algorithms or neural networks. In [14] the *Simulated Annealing* (SA) algorithm [15], has been proposed to minimize the co-channel interference in the uplink of two independent satellite systems. As a side remark, we point out that the problem of radio resource management (RRM) has been studied also in the context of terrestrial networks [16] [17], although the payload constraints and the different network topologies make the two optimization problems significantly different.

With respect to the state of the art, the present paper proposes a novel resource allocation strategy, in which a multi-objective optimization problem is addressed through the definition of an *ad-hoc* objective function. The optimization problem is addressed using a modified version of the SA algorithm. Unlike in [14], the present paper applies a variant of SA to the forward link of a multibeam satellite system adopting DVB-S2 technology and equipped with a flexible (in power and/or in bandwidth) payload. Furthermore, in our study we show the potentials of the proposed resource allocations scheme in presence of realistic requested traffic profile and operative conditions by considering different configurations of flexible payloads. Finally, the paper also attempts to shed some light on the most appealing payload design approach in terms of flexibility.

The remainder of this paper is structured as follows. Section II presents the system model and the formulation of the resource allocation problem, whereas the proposed allocation strategies are illustrated in Section III. Performance analysis and discussion of the results are provided in Section IV. Finally, Section V draws the conclusions of the investigation presented in this work and discusses some future research directions in the framework of radio resource management for next-generation satellite systems.

II. RADIO RESOURCE ALLOCATION PROBLEM

A. System Model

The present paper takes a multi-beam geostationary satellite system as reference. The satellite generates a geographical footprint subdivided into N_b beams, where each beam i , $i = 1, \dots, N_b$, serves N_u^i fixed satellite terminals. The population of users active on beam i generates an aggregate traffic request which we denote as T_r^i . Let us denote with $G_{i,j}$ the gain of the signal transmitted in beam i and received by user j , where j can take values between 1 and $N_u^{tot} = \sum_{i=1}^{N_b} N_u^i$, N_u^{tot} being

the total number of users in the system. Such gains account not only for the transmitted satellite antenna power, but also for the receiving and transmitting antennas gains and the propagation impairments (e.g., free space loss and atmospheric attenuation). Ideally, each satellite terminal j is expected to receive only the signal transmitted by its reference beam, which we denote as $\tilde{i}(j)$. However, due to the secondary lobes of the satellite antennas, user terminals suffer from interference generated by beams others than the reference one operating in the same frequency band (co-channel beams), leading to $G_{i,j} \neq 0$ for some $i \neq \tilde{i}(j)$. As far as the payload model is concerned, a single feed per beam (SFPB) architecture is considered. It is assumed that a number of TWTA equal to N_{TWTA} is available on-board the satellite and that each of them amplifies the same amount of bandwidth. Each tube has a total available DC power equal to P_{tot} . Each TWTA serves a subset of beams and reuses the whole bandwidth. The association between beams and the TWTA's is specified within the *connection matrix*. In the conventional system, the total bandwidth B of each TWTA is shared uniformly among the subset of amplified beams, so that the bandwidth per beam depends only on the specific coloring scheme adopted (e.g., B , $B/2$, or $B/4$ for 1, 2, or 4 colours, respectively). The TWTA bandwidth is shared among the connected beams in such a way that beams connected to the same tube cannot have overlapping portions of bandwidth. Data is transmitted through a beam making use of multiple carriers, each being assigned a fraction of the bandwidth allocated to the beam. The portions of data traffic addressed to the different users served by a given beam are multiplexed in time according to a time division multiplexing (TDM) framing.

B. Problem Formulation

Our aim is to allocate resources such that each beam receives an offered capacity T_o^i , $i \in \{1, \dots, N_b\}$, that is as close as possible to the requested capacity T_r^i while taking fairness into account. T_o^i depends on the bandwidth allocated to beam i and the power settings of the TWTA to which beam i is connected, as well as on the co-channel interference generated by other beams and on the channel gains (relative to both reference signal and interferers) of each single user. The optimization algorithm is run at the gateway. We assume that the gateway has knowledge of the gains $G_{i,j}$, $i \in \{1, \dots, N_b\}$, $j \in \{1, \dots, N_u^{tot}\}$, N_u^i being the number of users served by beam i . This assumption is a realistic one since fixed terminals are considered, for which the rate of channel variation can be assumed to be relatively slow. The channel gains are assumed to be periodically estimated by the gateway through a return channel. Such gains are used to choose the ModCod which is best suited to each terminal's current channel condition, i.e., the ModCod with the highest spectral efficiency that can be supported by the channel.

We consider three different payloads. The first payload can be optimized both in terms of bandwidth and power allocation. The second one has only bandwidth flexibility while in the third one only the TWTA operating conditions in terms of

power can be adjusted. For a fair comparison, both the number and the characteristics of the TWTA's are the same for all payloads. We also assume that the connection matrix, which determines the subset of beams that are connected to each TWTA, is the same in all payloads. The subset of beams connected to different TWTA's are disjoint, i.e., one beam cannot be connected to more than one TWTA. The flexibility in terms of bandwidth allows to modify in each TWTA the spectrum assignment to the subset of beams connected to it, under the constraint that the same portion of spectrum can not be assigned to more than one beam connected to the same TWTA. Note that the same portion of spectrum can be assigned from different TWTA's to one of the beams in their relative subset, so that from a resource allocation point of view each TWTA acts as an independent unit.

The flexibility in terms of power allocation consists in the possibility to change operating point (i.e., input back-off (IBO)) and power profile of each TWTA independently of the others. The available set of power profiles is $\{0, 1, 2, 3, 4\}$. The power profile indicates the attenuation (in dB) of the peak radio frequency power delivered by the TWTA with respect to the peak in the reference operation mode (i.e., power profile 0). A larger power profile indicates larger output back-off (OBO) for a given IBO, which reduces the non-linear effects of the tube at the cost of a reduced power efficiency. Payloads with no power flexibility keep a fixed IBO (equal to 3 dB) and a power profile equal to 2.

We fix the bandwidth granularity to 31.25 MHz, i.e., the bandwidth allocated to a beam must be a multiple of $B_{ch} = 31.25$ MHz. In the following we will refer to such elementary unit of bandwidth as *chunk*. The amount of bandwidth that can be assigned to a certain beam can be expressed as $N_{ch}B_{ch}$, where N_{ch} belongs to the set $\{1, 2, \dots, N_{ch}^{tot}\}$, N_{ch}^{tot} being the maximum number of chunks available in the system. In the present paper we assume an overall system bandwidth of 500 MHz, so that at most $N_{ch}^{tot} = 16$ chunks can be allocated to a beam. Transmission takes place on different carriers, each corresponding to a chunk. The optimization problem that we aim to solve is:

$$\begin{aligned} & \underset{\mathbf{v}, \mathbf{p}, \mathbf{B}}{\text{minimize}} && f(\mathbf{v}, \mathbf{p}, \mathbf{B}) \\ & \text{subject to} && v_t \in \{-20, -19, \dots, 5, 6\}, \quad t = 1, \dots, N_{\text{TWTA}} \\ & && p_t \in \{0, 1, 2, 3, 4\}, \quad t = 1, \dots, N_{\text{TWTA}} \\ & && \mathbf{B} \in \mathcal{B}, \end{aligned}$$

where N_{TWTA} is the number of TWTA's in the payload, $f(\cdot, \cdot)$ is the objective function to be minimized, which will be defined later on in this section, $\mathbf{v} = (v_1, \dots, v_{N_{\text{TWTA}}})$ and $\mathbf{p} = (p_1, \dots, p_{N_{\text{TWTA}}})$ are vectors containing the IBO's and the power profiles for all TWTA's, respectively, while \mathbf{B} is the $N_b \times N_{ch}$ bandwidth allocation matrix, which belongs to the set of feasible bandwidth allocation matrices \mathcal{B} . \mathcal{B} is a subset of the set of binary matrices whose structure depends on the specific payload bandwidth constraints.

1) *Key Performance Indicators*: As a common practice in optimization problems, we aim at minimizing an *objective*

function. The objective function reflects the system key performance indicators (KPI). In order to define the KPI we start with some general considerations. Our goal is to efficiently allocate the satellite resources with the aim of satisfying the requested traffic in all beams. The capacity request satisfaction can be looked at from a system (or global) perspective as well as from a beam (or user) perspective. From a global perspective, a valid choice would be to take as objective function a measure of the error in matching the requested capacity across the beams, i.e.,

$$E = \sum_{i=1}^{N_b} (T_o^i - T_r^i)^2.$$

Although this may be a valid indicator to measure the error with respect to an ideal resources allocation condition (by ideal we mean one in which the offered capacity matches the requested capacity exactly in each beam), the offered capacity exceeding the requested one is treated in the same way as the missing capacity, which is not desirable. Moreover, the measure E is potentially unbounded (i.e., it can assume arbitrary positive values) and this makes it difficult to evaluate the goodness of the optimization solution. Furthermore, even if relatively good results are obtained in terms of matching error E , it can still happen that traffic requests are largely unmatched for a non-negligible number of beams. This may be the case mainly for beams that present relatively low traffic request and for which too little resources are allocated. As a matter of fact, beams with relatively little capacity request may not have great impact on E and thus an optimization solution that performs well at a global scale may neglect such beams. Although beams with higher capacity requests are likely to be the most profitable ones from the satellite operator perspective, low traffic beams should be taken into account in the optimization process, since other factors may make such beams appealing (e.g., presence in the territory, reputation of the operator, etc.). Fairness is indeed a relevant parameter to be accounted for in the RRM optimization. Several different measures of fairness have been proposed in literature, such as the Jain Index and the normalized entropy. Using one of these measures as (negative) objective function would have the disadvantage of not accounting for the absolute value of the mismatch in terms of capacity (either excess or missing).

Let us now define the *satisfaction index* of beam i as $SI_i = \frac{T_o^i}{T_r^i}$. SI is a non-negative number which gives a measure of the extent up to which the requested capacity is satisfied. If $SI < 1$ the beam has been allocated insufficient resources for its capacity needs, while $SI > 1$ indicates that the beam is being over-provisioned. Ideally it would be good to keep track of both request satisfaction and absolute gap between the requested and the offered capacity. A way to visualize the system state in such terms can be the representation of all beams on a scatter plot in a plane having as axis the satisfaction index and the difference $\Delta_i = T_o^i - T_r^i$, which gives a measure of the missing (if negative) or wasted (if positive) capacity. We refer to such plane as the satisfaction/gap (SG)

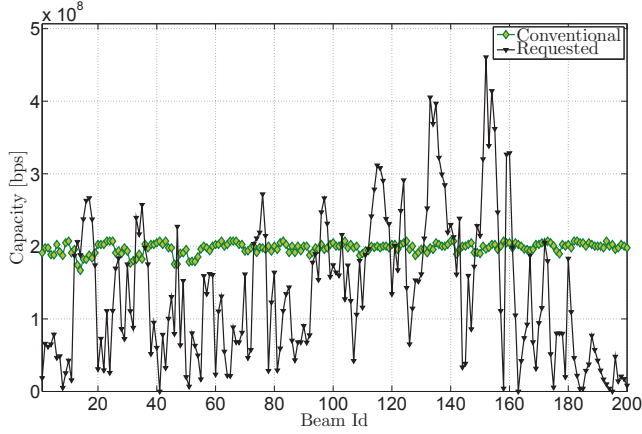


Fig. 1: Example of requested capacity and offered capacity for the conventional payload plotted versus beam Id.

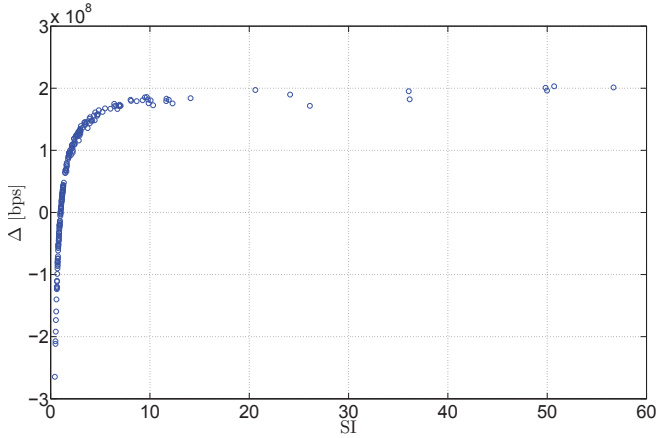


Fig. 2: Representation of beams in the SG plane. Each of the 200 beams is represented as a point (blue circle) in the plane having as x coordinate the satisfaction index and as y coordinate the capacity gap Δ as defined in this section.

plane. In Fig. 1 an example of requested and offered capacity for the conventional payload plotted versus the beam Id is shown. The corresponding SG representation is depicted in Fig. 2.

2) *Objective Function*: The plot in the SG plane gives a qualitative idea of the goodness of a given resource allocation solution in terms of both satisfaction and gap distribution. In order to have also a quantitative measure, we introduce a parameter which is derived from the SG plot. We will refer to it as the *satisfaction-gap measure* (SGM). The measure has been created so that the following hold:

- 1) Provide a measure of the mismatch with respect to the ideal case accounting for gap and satisfaction in all beams.
- 2) A beam with satisfaction lower than 1, say $1 - \delta$, has more weight with respect to a beam with satisfaction $1 + \delta$ (which is also undesired but not as bad as having beams with missing capacity).
- 3) Assume values in the interval $[0, 1]$, 1 being the desired

situation (perfect match of offered and requested capacity through all beams.¹).

The idea is to apply a transformation to the SG plane in such a way that the measure we look for satisfies the three conditions above. Let us start with point 1). In order to take both SI and Δ into account, we treat the SG plane as a complex plane, in which SI represents the real axis and Δ the imaginary axis. A beam/point is treated as a complex number in such plane. In order to satisfy point 2), we apply the following transformation to the beams with real part lower than 1:

$$\text{Re}\{c\} \rightarrow 1 - \frac{1}{\text{Re}\{c\}}, \quad \forall c : \text{Re}\{c\} < 1.$$

Note that after this transformation smaller SI translate to larger distances from the origin. In order to satisfy point 3) we shift the points with real part (satisfaction) larger than or equal to 1 towards the origin by applying the transformation $c \rightarrow c - 1$. In this way the point representing the optimal solution becomes $(0, 0)$. In Fig. 2 it can be seen how, depending on the unit of measure adopted to measure the excess/missing capacity (e.g., kbps, Mbps, Gbps) the range of the y axis can be quite wide with respect to the x axis. This can be easily fixed with a scaling operation. For each beam we do the following:

$$\text{Im}\{c\} \rightarrow \frac{\text{Im}\{c\}}{\beta},$$

with $\beta > 0$. The value of β can be chosen, for instance, equal to (or a function of) the system throughput of the conventional payload. In this way it is possible to make a comparison in terms of the goodness in the resource allocation solution between systems with different total capacities.

In order to get a measure which takes values between 0 and 1, we apply one last transformation to the plane which confines all the points within the circle of radius 1 around the origin. This is done applying the following transformation to the absolute value of each point, without modifying the phase of the number:

$$|c| \rightarrow 1 - e^{-|c|}. \quad (1)$$

According to this transformation, a point at infinite distance from the ideal condition (origin) will lie on the unitary circle after the transformation while a point that has $|c| = \delta \ll 1$ before applying the transformation in expression 1, will have a distance from the origin approximately equal to $1 - |c|$. The modified plot corresponding to the example in Fig. 2 is shown in Fig. 3.

Starting from the transformed plot, we define the SGM as measure of the average distance from the optimal condition:

$$\text{SGM} = 1 - \frac{1}{N_b} \sum_{i=1}^{N_b} |c_i|^3. \quad (2)$$

The third power rise in the sum on the right hand side of expression (2) is included in order to give more weight to

¹Note that such ideal condition can be achieved only in systems that are non-under-dimensioned in terms of total system bandwidth and power.

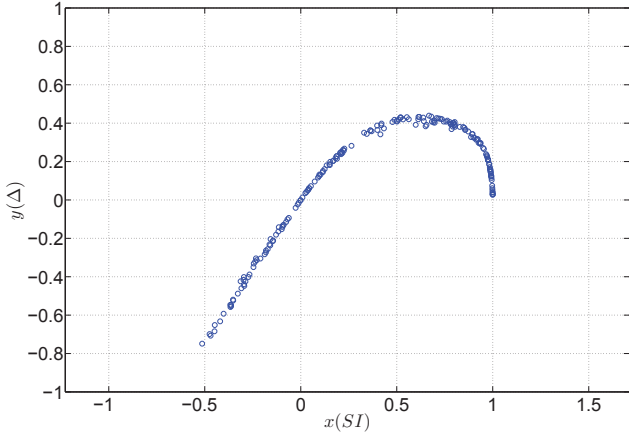


Fig. 3: Representation of beams in the modified SG plane. Each beam is represented as a point (blue circle) in the plane having as x coordinate a function of the satisfaction index and as y coordinate a function of the capacity gap Δ .

beams that are farther apart from the ideal condition (i.e., have SI which is either close to zero or much larger than 1 or have a large mismatch in terms of absolute capacity).

SGM is the complement to 1 of the average (cube of the) distance from the origin of the points in the transformed scatter plot. It can be easily seen that such measure takes values in $[0, 1]$ and is close to 1 when all points are gathered around the origin, which corresponds to the case in which the offered capacity matches almost exactly the requested capacity in each of the beams and there is little difference among the deltas.

The optimization problem to be solved is, finally:

$$\begin{aligned} & \underset{\mathbf{v}, \mathbf{p}, \mathbf{B}}{\text{minimize}} && -SGM(\mathbf{v}, \mathbf{p}, \mathbf{B}) \\ & \text{subject to} && v_t \in \{-20, -19, \dots, 5, 6\}, \quad t = 1, \dots, N_{\text{TWTA}} \\ & && p_t \in \{0, 1, 2, 3, 4\}, \quad t = 1, \dots, N_{\text{TWTA}} \\ & && \mathbf{B} \in \mathcal{B}. \end{aligned}$$

III. RESOURCE ALLOCATION STRATEGY

Even assuming full channel state information at the transmitter, finding the optimal resource allocation is not trivial. This is due, on one side, to the non-convexity of the objective function and on the other side to the large number of possible solutions, which makes exhaustive search not viable².

We propose a suboptimal algorithm based on a slightly modified version of the Simulated Annealing algorithm [15]. The algorithm tries to minimize the objective function defined in Section II-B. This is done by running iteratively the SA algorithm, each time using lower starting and stopping temperatures. The way the SA algorithm is applied at each run is described in the following.

A. Perturbation of the Feasible Point

The SA algorithm uses as starting point the same bandwidth and power allocation as a conventional payload.

²For a flexible payload with 50 beams, 8 bandwidth chunks and 20 allowed IBO levels, the number of possible allocations (feasible points) is equal to $(256 \times 20)^{50}$ which is on the order of 10^{185}

At each iteration the algorithm perturbs the feasible point. Depending on the payload to which the algorithm is applied, either the bandwidth, the power or both can be modified. For the payload with full flexibility the algorithm chooses randomly at each iteration whether to modify one of the other.

The perturbation of the feasible point is done as follows. A beam is selected at random, then:

- If the bandwidth is to be modified, the number of bandwidth chunks N_{ch} currently allocated to the beam is modified by adding to such number a random variable $u \in \{-1, 0, +1\}$ while keeping the number of allocated beams within the set $\{1, 2, \dots, N_{ch}^{tot} - 1, N_{ch}^{tot}\}$. Once the new number of chunks N_{ch} is selected, their location in the bandwidth is selected at random among the $\binom{N_{ch}^{tot}}{N_{ch}}$ possible dispositions. Afterwards, the algorithms switches off the chunks allocated to the selected beam from the other beams connected to the same TWTA (if necessary).
- If power is to be modified, the algorithm selects the TWTA to which the selected beam is connected and modifies either its IBO or its power profile. Modifying the operating conditions of the TWTA induces a modification in the amount of power delivered by the TWTA, its power efficiency (i.e., the ratio of the delivered RF power to the absorbed DC power) and the intermodulation interference associated with the TWTA nonlinearity. All these effects are taken into account by the algorithm through realistic payload models. Note also that all the beams connected to the same TWTA are affected by the same attenuation/amplification of the signal on the selected beam. This is done in order to avoid the so-called *capture effect*, which takes place in TWTA's when carriers of different power are fed to the amplifier [18].

B. SGM Evaluation

Once the resources of the selected beam and the corresponding TWTA have been modified, the resulting SINR for all users are calculated for each bandwidth chunk³. The new SINR's are used to determine the ModCod with highest spectral efficiency supported by the channel of each terminal in each chunk according to the DVB-S2 standard. Once the spectral efficiency for all terminals and all allocated chunks are obtained, they are averaged out across users and chunks. More specifically, let us consider a specific beam i . We call $\eta_{j,c}^i$ the spectral efficiency achievable by terminal j in chunk c of beam i . The average spectral efficiency in beam i is then:

$$\bar{\eta}^i = \frac{1}{N_u^i N_{ch}^i} \sum_{j=1}^{N_u^i} \sum_{c=1}^{N_{ch}^i} \eta_{j,c}^i, \quad (3)$$

where N_u^i is the number of users in beam i while N_{ch}^i is the number of chunks allocated to beam i . (3) follows from the assumption that all users within a beam access their content in a TDM fashion on all chunks, such that each chunk is

³The SINR is calculated on a chunk-by-chunk basis since each chunk is assumed to be a single carrier.

allocated to a given user during the whole assigned reception slot. Finally, using the average spectral efficiency together with the number of bandwidth chunks allocated to each beam (taking the roll-off α into account), the offered capacity for beam i is calculated as:

$$T_o^i = \frac{1}{1 + \alpha} B_{ch} N_{ch}^i \bar{\eta}^i.$$

The T_o^i of all beams are then used to compute the new value of the SGM as described in Section II-B.

C. Feasible Point Update

Once the new SGM is calculated, there are two possibilities:

- The SGM obtained in the new point is larger than the old one. In this case the new point is kept and a new iteration starts.
- The SGM obtained in the new point is smaller than the old one. In this case the new point is kept with a certain probability. The probability of keeping the new point depends on a simulation parameter that is updated periodically. As common practice in the literature related to simulation annealing, such parameter is called *temperature* and is indicated in the following as T_{sa} . The new point is accepted with probability:

$$e^{-\left(\frac{SGM_{old} - SGM_{new}}{SGM_{old}}\right) T_{sa}}.$$

Since we are considering the case in which the new point is worse than the previous one, in the expression above we always have $SGM_{old} - SGM_{new} > 0$. Note also that such probability decreases as the temperature T_{sa} decreases.

The temperature is decreased once a predefined number of iterations is reached. The cooling law at iteration n is:

$$T_{sa}(n) = \Delta T \times T_{sa}(n-1), \quad (4)$$

where $0 < \Delta T < 1$. The block diagram describing one call of the algorithm is shown in Fig. 4.

As mentioned earlier in this section, the proposed algorithm is a modified version of the SA. The modification consists in that the SA algorithm is run iteratively, each time using lower starting and stopping temperatures. Specifically, if we indicate with $T_start[l]$ and $T_stop[l]$ the starting and stopping temperatures at the l -th algorithm call, respectively, the following holds:

$$T_start[l] = T_stop[l-1].$$

The reason behind such modification is that, for the specific problem considered, we observed a tendency of the SA to converge to local minima. This is a well known behavior of stochastic optimization algorithm with non-convex objective functions. The typical solution usually adopted is to run the algorithm more than once, each time starting from a different starting point. In the setup we study such solution showed limited advantages. For this reason we introduced a variant of such approach in which i) the starting point of the new

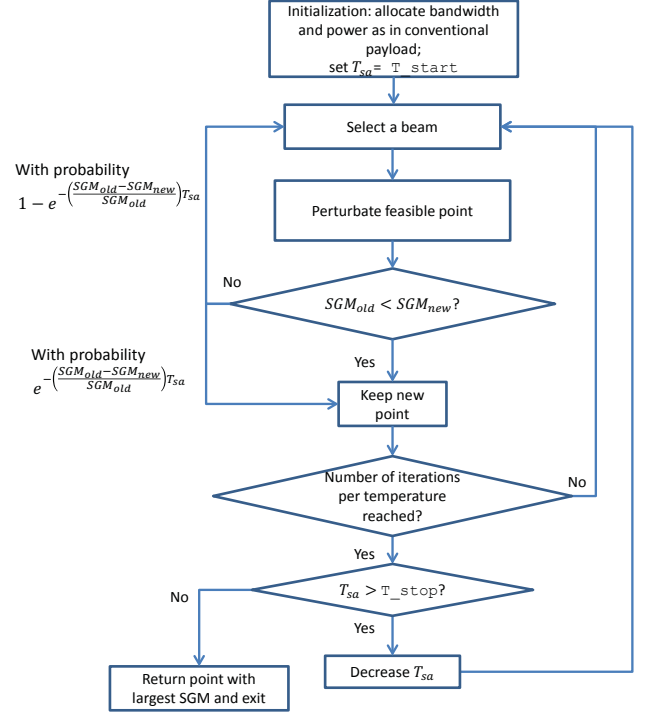


Fig. 4: Flow diagram for one call of the SA algorithm. At each call the initial and starting temperatures T_start and T_stop are decreased such that $T_start[l] = T_stop[l-1]$, l being the call index.

run is the feasible point output of the previous one, ii) rather than starting the simulator anew, we decrease the starting and stopping temperature at each call. The overall effect is to break the path of the algorithm in the feasible set, avoiding that it gets stuck in regions characterized by values of SGM that are lower than the last accepted one.

IV. PERFORMANCE ANALYSIS

In the following we present the results obtained by applying the proposed algorithm to three different payloads. The first payload we consider has full flexibility, in the sense that it can be optimized both in terms of bandwidth and power allocation. The second payload has only bandwidth flexibility, while in the third one only the power setting can be modified. The starting point of the algorithm is the resource allocation used in the conventional payload. In all simulations we fixed $N_{TWT A} = 50$, $B = 500$ MHz, $B_{ch} = 31.25$ MHz, $N_b = 200$ and $N_u^{tot} = 2000$ (with $N_u^i = 10$, $\forall i = 1, \dots, N_b$).

We compare the results for the different payloads in terms of both the effectiveness of the algorithm to meet the requested capacity and the fairness with which the different beams are treated. Specifically, the Jain Index of the ceiled satisfaction index is used to measure the fairness in the system. The ceiled satisfaction index is defined as $\bar{SI} = \min\{SI, 1\}$ and is a measure of the satisfaction level of a beam which focuses on the missing capacity. The JI is calculated as:

$$JI = \frac{\left(\sum_{i=1}^{N_b} \bar{SI}_i\right)^2}{N_b \sum_{i=1}^{N_b} \bar{SI}_i^2}. \quad (5)$$

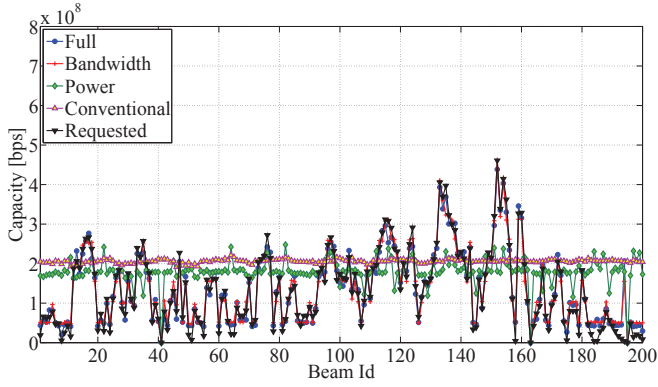


Fig. 5: Capacity versus beam number. The requested capacity at 00:00 h is shown together with the capacity offered by the conventional payload and the three flexible payloads considered.

Another relevant figure of merit for satellite communications systems is the *unmet capacity* (UC), which is the overall amount of requested capacity that can not be met. UC is defined as:

$$UC = \sum_{i=1}^{N_b} (T_r^i - T_o^i)^+, \quad (6)$$

where $(x)^+ = \max(x, 0)$. Similarly as for UC, we define the *excess capacity* as:

$$EC = \sum_{i=1}^{N_b} (T_o^i - T_r^i)^+, \quad (7)$$

which is the sum across the beams of the offered capacity exceeding the requested capacity. The UC and the EC give an indication of the effectiveness of the resource allocation. Finally, we define the *total offered capacity* (TOC) as:

$$TOC = \sum_{i=1}^{N_b} T_o^i. \quad (8)$$

In figures 5 and 6 the requested capacity is plotted against the beam Id together with the offered capacity obtained by applying the proposed algorithm to different payloads. Specifically, the capacity offered by the payload with both bandwidth and power flexibility (Full), the one with only bandwidth flexibility (Bandwidth) and the one with power flexibility only (Power) obtained with the proposed algorithm are shown. The offered capacity of the conventional payload is shown as a benchmark. The requested capacity has been generated according to a traffic model developed by the German Aerospace Center (DLR), accounting for the geographical and time variations of traffic requests as well as the availability of satellite network. The model provides good matches with real requested traffic statistics, as discussed in [5, section III-E], to which the interested reader can refer for more details. In Table I and Table II the comparison among the four different payloads is presented for the requested traffic at off peak (00:00) and peak (19:00) hours, respectively. In order to have a deeper understanding of the SGM as performance metric and the

TABLE I: SGM, Jain Index, unmet capacity and excess capacity at off-peak hour (00:00) for the four payloads. Values are rounded to the third decimal.

	SGM	Jl	UC [Gbps]	EC [Gbps]	TOC [Gbps]
Full	0.923	0.995	1.34	1.664	27.222
Bandw.	0.905	0.995	1.33	1.984	27.552
Power	0.62	0.978	4.263	13.26	35.895
Conv.	0.567	0.982	3.237	17.509	39.529

implications of using it as objective function, in the tables SGM, Jain Index, unmet capacity and excess capacity are shown for each payload. All values are rounded to the third decimal. Since the last three parameters have been previously used in literature or have an intuitive interpretation, they help to understand the SGM more in depth.

With reference to Fig. 5, the payload with bandwidth flexibility and the payload with full flexibility are able to provide an offered capacity which closely follows the requested one in most of the beams. In beams with very low requested capacity, such as beam with Id 194, the offered capacity is relatively larger than the requested one. This is in part due to the limited granularity in terms of bandwidth. The payload with power flexibility is not able to follow the requested traffic as closely as the other two flexible payloads. One of the reasons for this is the fact that the power can be optimized only at TWTA level, so that all beams experience the same power increase or decrease, while this is not the case for the payload with bandwidth flexibility, in which the number of bandwidth chunks assigned to a beam can be different from that of other beams connected to the same TWTA (provided the same chunk of bandwidth is not allocated to more than one of the beams connected to it). As a last remark, we notice that the offered capacity of the conventional payload shows some fluctuations across the beams. These are due to the slight differences in channel gains and interference levels experienced by the different users, that, on turn, depend on the realistic satellite antenna radiation pattern used. From Table I we can see that the qualitative considerations presented above are backed up by the numerical values in the table. As a matter of fact, the SGM is higher in the payloads with full and bandwidth flexibility. This corresponds to higher Jl, lower UC and lower EC with respect to the other two payloads. Comparing the Full and the Bandwidth payloads we see that the UC in the two payloads are almost the same while the smallest EC is achieved by the Full payload, which leads to a larger SGM. This shows how the SGM jointly accounts for UC, EC and fairness, although the mapping from SGM to the three measures is not straightforward.

In Fig. 6 we show the results relative to a more demanding pattern of requested capacity. The setup is much more challenging than the one presented in Fig. 5 from an optimization perspective. This can be inferred from the peaks of requested traffic reaching more than three times the capacity offered by the conventional payload, and from the fact that the overall system bandwidth and TWTA number in all the advanced payloads is the same as in the conventional one

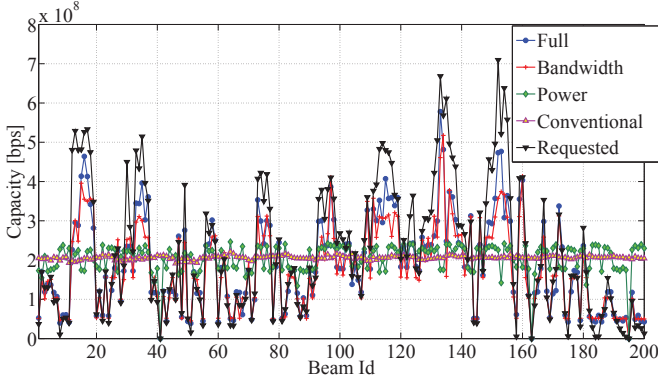


Fig. 6: Capacity versus beam number. The requested capacity at 19:00 h is shown together with the capacity offered by the conventional payload and the three flexible payloads considered.

TABLE II: SGM, Jain Index, unmet capacity and excess capacity at peak hour (19:00) for the four payloads. Values are rounded to the third decimal.

	SGM	Jl	UC [Gbps]	EC [Gbps]	TOC [Gbps]
Full	0.912	0.978	8.514	1.161	37.415
Bandw.	0.884	0.966	10.692	1.081	35.157
Power	0.638	0.93	14.808	10.779	40.738
Conv.	0.603	0.914	15.765	12.166	39.529

in both simulations. Also in this case the payloads with full and bandwidth flexibility can best follow the requested traffic profile, while the payload with power flexibility has only a limited adaptation capability. Interestingly, even though the power flexibility alone is not able to follow the requested traffic, it provides an advantage to the payload with full flexibility. This can be seen from the fact that the Full payload can better approximate the highest peaks of requested capacity with respect to the bandwidth flexible one. This is further confirmed by the results shown in the Table II. In the table we see that the payload with full flexibility achieves the best performance in all the four figures of merit considered, reducing the unmet and the excess capacity and increasing the system fairness with respect to any of the other payloads. In this case we see how a higher SGM corresponds to a better performance through the whole spectrum of figures of merits considered.

V. CONCLUSIONS

We studied the problem of radio resource management in multibeam satellite systems. A novel objective function has been introduced with the aim to match the requested capacity across the beams as close as possible while taking fairness into account. We proposed a stochastic optimization algorithm to minimize such function based on a modified version of the simulated annealing algorithm. We applied the algorithm to three payloads having different degrees of flexibility, namely flexibility both in bandwidth and power, in bandwidth only and in power only. Realistic payload models, antenna pattern, co-channel interference and requested traffic distribution were used in the simulations. Our results show that the proposed approach is much more efficient than the traditional conventional

payload in matching the requested capacity across the beams and leads to interesting results both under low and high traffic demand. The goodness of the proposed approach has been supported by measuring different figures of merit traditionally used in this kind of analysis, namely missing capacity, excess capacity and Jain index.

REFERENCES

- [1] D. Serrano-Velarde, E. Lance, and H. Fenech, "Novel dimensioning method for high-throughput satellites: forward link," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 2146–2163, July 2014.
- [2] H. Fenech, A. Tomatis, S. Amos, V. Soumphonphakdy, and J. L. S. Merino, "Eutelsat HTS systems," *Int. J. of Satellite Commun. and Networking*, pp. n/a–n/a, 2016. [Online]. Available: <http://dx.doi.org/10.1002/sat.1171>
- [3] ETSI, "DVB, User Guidelines for the Second Generation System for Broadcasting, Interactive Services, News Gathering and Other Broad-band Satellite Applications (DVB-S2)," Technical Report, ETSI, Tech. Rep., Feb. 2005, eTSI TR 102 376 V1.1.1.
- [4] M. A. V. Castro and G. S. Granados, "Cross-layer packet scheduler design of a multibeam broadband satellite system with adaptive coding and modulation," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 1, pp. 248–258, Jan. 2007.
- [5] X. Alberti, J. Cebrian, A. D. Bianco, Z. Katona, J. Lei, M. Vazquez-Castro, A. Zanus, L. Gilbert, and N. Alagha, "System capacity optimization in time and frequency for multibeam multi-media satellite systems," in *Advanced Satellite Multimedia Systems Conf.*, Cagliari, Italy, Sep. 2010, pp. 226–233.
- [6] J. Lizarraga, P. Angeletti, N. Alagha, and M. Aloisio, "Flexibility performance in advanced ka-band multibeam satellites," in *IEEE Int. Vacuum Electronics Conf.*, Apr. 2014, pp. 45–46.
- [7] J. Lei and M. A. Vazquez-Castro, "Joint power and carrier allocation for the multibeam satellite downlink with individual sinr constraints," in *IEEE Int. Conf. on Commun.*, May 2010, pp. 1–5.
- [8] J. Lei and M. A. Vazquez-Castro, "Multibeam satellite frequency/time duality study and capacity optimization," *J. of Commun. and Networks*, vol. 13, no. 5, pp. 472–480, Oct 2011.
- [9] J. P. Choi and V. W. S. Chan, "Optimum power and beam allocation based on traffic demands and channel conditions over satellite downlinks," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 6, pp. 2983–2993, Nov. 2005.
- [10] —, "Resource management for advanced transmission antenna satellites," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 3, pp. 1308–1321, Mar. 2009.
- [11] T. Qi and Y. Wang, "Energy-efficient power allocation over multibeam satellite downlinks with imperfect CSI," in *Int. Conf. on Wireless Commun. Signal Processing*, Oct. 2015, pp. 1–5.
- [12] A. Aravanis, B. Shankar, M. R., P. Arapoglou, G. Danoy, P. Cottis, and B. Ottersten, "Power allocation in multibeam satellite systems: A two-stage multi-objective optimization," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 6, pp. 3171–3182, June 2015.
- [13] B. G. Evans, P. T. Thompson, and A. Kyrgiazos, "Irregular beam sizes and non-uniform bandwidth allocation in hts," in *AIAA Int. Commun. Satellite Systems Conf.*, Florence, Italy, Sep. 2013, pp. 1–7.
- [14] S. Salcedo-Sanz, R. Santiago-Mozos, and C. Bousio-Calzón, "A hybrid Hopfield network-simulated annealing approach for frequency assignment in satellite communications systems," *IEEE Trans. on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 2, pp. 1108–1116, Apr. 2004.
- [15] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.
- [16] T. Olwal, K. Djouani, and A. Kurien, "A survey of resource management towards 5G radio access networks," *IEEE Commun. Surveys Tutorials*, vol. PP, no. 99, pp. 1–1, 2016.
- [17] H. Zhang, C. Jiang, N. Beaulieu, X. Chu, X. Wang, and T. Quek, "Resource allocation for cognitive small cell networks: A cooperative bargaining game theoretic approach," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 6, pp. 3481–3493, June 2015.
- [18] G. Maral and M. Bousquet, *Satellite Communications Systems: Systems, Techniques and Technology*, 5th ed. John Wiley and Sons Ltd, 2009.